The ISME Journal (2010) 4, 882–895

www.nature.com/ismej

ORIGINAL ARTICLE

Fine-scale evolution: genomic, phenotypic and ecological differentiation in two coexisting *Salinibacter ruber* strains

Arantxa Peña[1], Hanno Teeling[2], Jaime Huerta-Cepas[3,4,10], Fernando Santos[1], Pablo Yarza[5], Jocelyn Brito-Echeverría[5], Marianna Lucio[6], Philippe Schmitt-Kopplin[6], Inmaculada Meseguer[7], Chantal Schenowitz[8], Carole Dossat[8], Valerie Barbe[8], Joaquín Dopazo[3,4], Ramon Rosselló-Mora[5], Margarete Schüler[2], Frank Oliver Glöckner[2,9], Rudolf Amann[2], Toni Gabaldón[3,10] and Josefa Antón[1]

[1]*Departamento de Fisiología, Genética y Microbiología, and IMEM, Universidad de Alicante, Apartado 99, Alicante, Spain;* [2]*Max Planck Institute for Marine Microbiology, Celsiusstrasse 1, Bremen, Germany;* [3]*Department of Bioinformatics, Centro de Investigación Príncipe Felipe, Autopista del Saler 16, Valencia, Spain;* [4]*Functional Genomics Node (INB), Centro de Investigación Príncipe Felipe (CIPF), Valencia, Spain;* [5]*Marine Microbiology Group, Institut Mediterrani d'Estudis Avançats, Esporles, Spain;* [6]*Helmholtz Zentrum Muenchen, German Research Center for Environmental Health, Institute of Ecological Chemistry, Neuherberg, Germany;* [7]*Departamento de Producción Vegetal y Microbiología, Universidad Miguel Hernández. Avda. de la Universidad s/n., Elche (Alicante), Spain;* [8]*CEA/DSV/IG/Genoscope, Evry, France and* [9]*Jacobs University Bremen gGmbH, Campusring 1, Bremen, Germany*

Genomic and metagenomic data indicate a high degree of genomic variation within microbial populations, although the ecological and evolutive meaning of this microdiversity remains unknown. Microevolution analyses, including genomic and experimental approaches, are so far very scarce for non-pathogenic bacteria. In this study, we compare the genomes, metabolomes and selected ecological traits of the strains M8 and M31 of the hyperhalophilic bacterium *Salinibacter ruber* that contain ribosomal RNA (rRNA) gene and intergenic regions that are identical in sequence and were simultaneously isolated from a Mediterranean solar saltern. Comparative analyses indicate that *S. ruber* genomes present a mosaic structure with conserved and hypervariable regions (HVRs). The HVRs or genomic islands, are enriched in transposases, genes related to surface properties, strain-specific genes and highly divergent orthologous. However, the many indels outside the HVRs indicate that genome plasticity extends beyond them. Overall, 10% of the genes encoded in the M8 genome are absent from M31 and could stem from recent acquisitions. *S. ruber* genomes also harbor 34 genes located outside HVRs that are transcribed during standard growth and probably derive from lateral gene transfers with *Archaea* preceding the M8/M31 divergence. Metabolomic analyses, phage susceptibility and competition experiments indicate that these genomic differences cannot be considered neutral from an ecological perspective. The results point to the avoidance of competition by micro-niche adaptation and response to viral predation as putative major forces that drive microevolution within these *Salinibacter* strains. In addition, this work highlights the extent of bacterial functional diversity and environmental adaptation, beyond the resolution of the 16S rRNA and internal transcribed spacers regions.

The ISME Journal (2010) **4,** 882–895; doi:10.1038/ismej.2010.6; published online 18 February 2010
Subject Category: evolutionary genetics
Keywords: comparative genomics; microdiversity; micro-evolution; phage; *Salinibacter*

Introduction

Microbial genomics has undergone an amazing development since the finishing of the first bacterial genome in 1995 (Fleischmann *et al.*, 1995). Although pathogens and extremophiles with biotechnological potential were initially at focus, recently attention has shifted to environmentally relevant micro-organisms that constitute the majority of the existing microbial biodiversity. Marine habitats, covering

Correspondence: J Antón, Departamento de Fisiología, Genética y Microbiología, Universidad de Alicante, División de Microbiología, Alicante 03080, Spain.
E-mail: anton@ua.es
[10]Current address: Bioinformatics and Genomics Programme, Center for Genomic Regulation (CRG), Dr Aiguader, 88, Barcelona 08003, Spain.
Received 28 October 2009; revised 4 January 2010; accepted 5 January 2010; published online 18 February 2010

more than 70% of our planet's surface, have gained special attention in microbial genome sequencing, as illustrated by more than 200 available draft genomes contributed by the Marine Microbiology Initiative of the Gordon and Betty Moore foundation (http://www.moore.org).

The possibility of comparing complete genomes of related prokaryotes provides unique opportunities to study the course of evolution at the DNA level. Although comparisons of genomes of different species give insights on a larger scale of evolution, the so-called microevolution (Abby and Daubin, 2007) can only be addressed by comparing genomes of different strains within the same species. Microevolution studies can reveal how bacterial genomes evolve, provide insights in the hidden diversity within a species' population and elucidate the ecological mechanisms that drive microbial radiation (Sikorski, 2008). Current knowledge on microevolution stems mostly from studying strains of pathogenic bacteria but is unmatched for environmentally relevant prokaryotes. So far, comparative studies on genomic adaptations of environmentally relevant prokaryotes have focused on ecotypes, that is, populations of the same species with distinct ecological strategies (Cohan and Koeppel, 2008). Examples are studies on *Prochlorococcus* sp. (Coleman *et al.*, 2006) or *Alteromonas macleodii* (Ivars-Martínez *et al.*, 2008) isolates, which have ecotypes that are also distinguishable through their 16S ribosomal RNA (rRNA) gene(s) or 16S-23S rRNA internal transcribed spacers (ITS). These ecotypes, however, would not fall within the 16S rRNA gene microdiversity clusters (>99% similarity) described by Acinas *et al.* (2004) for coastal bacterioplankton. These clusters have been subsequently described for other environments, including hypersaline systems, and have been corroborated by metagenomic data (Wilmes *et al.*, 2008). The widespread presence of this microdiversity within microbial populations raises three main questions: (i) How much functional and genomic diversity is hidden behind this diversity? (ii) What is the ecological meaning of this microdiversity? and (iii) How does this microdiversity arise?

In few cases these questions have been answered. For instance, *Vibrio splendidus*, which is composed of strains that are 99% identical in 16S rRNA gene sequences and shows a wide genomic diversity (Thomson *et al.*, 2005), differentiates into 15 microdiverse habitat-associated clusters, possibly by invading new niches or partitioning resources at increasingly fine scales (Hunt *et al.*, 2008). In other cases, response to viral predation and neutral diversification have been suggested as mechanisms for generating microdiversity (Wilmes *et al.*, 2008). In any case, comparisons of both genomes and putative ecological traits for very closely related strains are so far very scare (for examples, see Wilmes *et al.* (2008) and below in this paper).

The genus *Salinibacter* can be found in hypersaline environments worldwide. *Salinibacter* populations have been detected by culture and molecular approaches in crystallizer ponds and salt lakes in places as distant as Australia, the Peruvian Andes, Turkey and Spain (Antón *et al.*, 2008). These hypersaline environments harbor dense communities of extremely halophilic *Archaea*, often dominated by the square-shaped archaeon *Haloquadratum walsbyi*. Nevertheless, *Salinibacter ruber* can account for a large part of the respective microbial communities, with population sizes ranging up to more than 10^7 cells ml^{-1} (around 30% of total cell counts). The description of the species *S. ruber* was based on five strains (Antón *et al.*, 2002), of which two (M8 and M31, see below) were simultaneously isolated in September 1999 from two samples taken from the same crystallizer pond of a coastal solar saltern in Mallorca (Balearic Islands, Spain). A multi-locus sequencing analysis among 10 *S. ruber* strains, including M8 and M31, from salterns in three distant locations (Atlantic, Andean and Mediterranean) could not show a genetic divergence related to the geographical origin of the isolates (Rosselló-Mora *et al.*, 2008). Phylogenetic reconstructions using these selected genes resulted in different tree topologies, a phenomenon that has been previously observed in extremely halophilic *Archaea* from solar salterns (Papke *et al.*, 2004) and is considered to be an indication for high intra-specific recombination rates (Feil, 2004). Consistent with this, a split decomposition (Huson and Bryant, 2006) of the concatenated sequences of the genes provided statistically significant evidence for recombination among these 10 *S. ruber* strains (Antón *et al.,* unpublished results).

Strains M8 and M31 share an identical rRNA operon sequence, including identical ITS. Among all the isolated *S. ruber* strains M8 and M31 are among the most closely related (based on rRNA, Pulsed Fiel Gel Electroforesis and Random **A**mplification of Polymorphic DNA analysis) (Peña *et al.*, 2005; Antón *et al.*, 2008), and would thus represent an example of very short-scale microdiversity within this species. However, their level of DNA–DNA hybridization is around 88%, which indicates that they have noteworthy genomic differences (Peña *et al.*, 2005). These differences are intriguing given that they inhabit a rather homogenous environment that at first glance offers little opportunities for micro-niche differentiations.

The key questions are: what are the genomic differences among these strains?, how did they arise and to which extent are they neutral or have an ecological relevance? For this purpose, the genome of strain M8 has been fully sequenced and compared with the previously sequenced M31 genome (type strain DSMZ 13855^T, Mongodin *et al.*, 2005). Phenotypic differences between both strains were examined by metabolomic analysis using Fourier transform ion cyclotron resonance mass spectrometry (Rosselló-Mora *et al.*, 2008). To explore their

putative ecological differences in nature, their phage susceptibility against different saltern waters was assayed, and competition experiments between both strains carried out. In terms of microevolution, this is the shortest-scale study on genomic changes and adaptations with representatives of the same species, beyond the ecotype level, conducted so far.

Materials and methods

Architecture and analysis of M8 genome sequence
Potential protein coding open reading frames were identified with mORFind v2 (Waldmann and Teeling, unpublished) (Supplementay Material). The sequences have been submitted to EMBL database (http://www.ebi.ac.uk/embl) (accession nos. FP565814 (chromosome) and FP565810 (plasmid pSR11), FP565811 (pSR56), FP565812 (pSR61) and FP565813 (pSR84)).

Phylome reconstruction and orthology determination
We reconstructed a complete set of phylogenetic trees, also known as phylome, for the two *S. ruber* genomes. All trees and aligments can be accessed through www.phylomedb.org (Huerta-Cepas *et al*., 2008) (Supplementary Figure 1). Orthologous and paralogous relationships among genes encoded in the two *Salinibacter* genomes were determined through a phylogenetic approach. (Supplementay Material).

Identification of genes putatively involved in Bacteria-Archaea *interdomain lateral gene transfer (LGT)*
We adopted a strategy of combining phylogeny and oligonucleotide frequency analyses to identify genes putatively involved in interdomain LGT events (Supplementay Material).

Metabolomic analysis
Metabolite profiles for extracellular and pellet fractions of pure cultures of M8 and M31 were analyzed by Fourier transform ion cyclotron resonance mass spectrometry (FT-ICMS) (Supplementay Material).

Competition experiments between M8 and M31 and phage susceptibility
Co-cultures of M8 and M31 were monitored by real-time PCR using strain-specific primers (Supplementay Material). Phage susceptibility for each strain was monitored by the plaque-assay technique as described in Supplementay Material.

Results and discussion

Genomic differences among M8 and M31 strains: plasticity and mosaic structure
Overall comparison between M8 and M31 genomes. The complete genome sequence of strain M8 was determined, analyzed and compared with the previously published M31 genome (Mongodin *et al*., 2005). The main results from these analyses are summarized in Table 1 and Figures 1 and 2.

With the exception of hypervariable regions or genomic islands (HVRs, described in detail below), high levels of conservation in gene order and orientation could be observed between M8 and M31 chromosomes (see dot-plot in Supplementary Figure 2). Such degree of synteny is expected because of the phylogenetic proximity of the two strains. However, 518 (18%) of the M8/M31 orthologous gene pairs had rather low levels of amino acid sequence identities, varying on a scale from 23% to 90%. In general, these divergent orthologs were evenly distributed throughout the whole genome (Figure 2, fourth circle inward; Figure 1), although the most divergent genes were concentrated in HVRs (see below).

Overall, M8 and M31 share around 90% of their genes or, in other words, there are 10% strain-specific genes. Among the shared genes, 17% had no homologs in public repositories and should thus for now be considered as species specific. The species-specific genes would be part of the *S. ruber* core genome (Feil, 2004), while the strain-specific genes are part of the species' accessory genome (Figure 3). Together, core and accessory genomes will be part of the species pan-genome (Meddini *et al*., 2005). Although most strain-specific genes were concentrated in HVRs, many gene insertions and/or deletions and transpositions could be observed also in aligned (that is, non-HVR) areas of both chromosomes, which, together with the 'divergent' genes (see above) accounted for a considerable part of the differences outside the HVRs. The remarkable presence of indels outside hypervariable regions has been previously described for *Pelagibacter ubique* isolates. Whilhelm *et al*. (2007) compared the genomes of two *P. ubique* isolates from the Oregon coast (one base difference in their 16S rRNA genes; 98% ITS identity; genome size of approximately 1.3 Mb) and found 62 gene indels outside HVRs. The investigators suggested that these indels could be polymorphisms, similar to allelic polymorphisms. The number of such indels in *S. ruber* M8 is 166, which is similar to that of *P. ubique* when normalized for the chromosome size (45.9 versus 47.7 Mb^{-1}).

The identified orthologous gene pairs exhibited an average amino acid identity of 94.2% and an average nucleotide identity (ANI) of 93.5%. This ANI value is lower than that derived from random fragment blast pairs (ANIb, 98.5%), which can be explained by the different thresholds used in their calculations. In this regard, the level of nucleotide identity between orthologous pairs reached 98.4 when indels were not considered (exact matches over aligned residues), a level that is more comparable to that of ANIb. It is also reasonable to assume that an ANI calculated with orthologs reflects existing evolutionary differences more accurately

Table 1 Comparison of *Salinibacter ruber* DSM 13588 general features with *S. ruber* strain M8

	S. ruber			
	M8		*M31 DSM 13588*	
Chromosome				
Length (pb)	3 619 447		3 551 823	
G+C content	66.12%		66.29%	
ORF number	3086		2934	
rRNAs	3		3	
tRNAs	43		44	
Total aligned sites:	3 348 702		3 283 845	
% Total ANI	98.45		nd	
% similarity (aa)		94.3		
% similarity (nt)		93.5		
Total orthologous[a]		2876		
COG categories[b]	*No*	*% of total*	*No*	*% of total*
J-Translation, ribosomal structure and biogenesis	134	4.34	106	3.69
K-Transcription	106	3.43	110	3.83
L-DNA replication, recombination and repair	165	5.34	145	5.05
D-Cell division and chromosome partition	21	0.68	23	0.80
V-Defense mechanisms	36	1.17	47	1.64
O-Posttranslational modification, protein turnover, chaperones	94	3.04	85	2.96
M-Cell envelope biogenesis, outer membrane	159	5.15	147	5.12
N-Cell motility and secretion	54	1.75	57	1.99
U-Intracellular trafficking, secretion, and vesicular transport	49	1.59	51	1.78
T-Signal transduction mechanisms	139	4.50	141	4.91
C-Energy production and conversion	132	4.30	123	4.29
G-Carbohydrate transport and metabolism	106	3.43	86	3.00
E-Amino acid transport and metabolism	205	6.64	182	6.34
P-Inorganic ion transport and metabolism	127	4.11	117	4.08
F-Nucleotide transport and metabolism	69	2.24	60	2.10
H-Coenzyme metabolism	109	3.53	83	2.89
I-Lipid metabolism	83	2.69	72	2.51
Q-Secondary metabolites biosynthesis, transport and catabolism	64	2.07	53	1.85
R-General function prediction only	274	8.88	257	8.96
S-Function unknown	164	5.31	132	4.60
No COG categories	1090	35.32 %	989	34.46 %
Conserved hypothetical protein	1005	32.57	320	10.91
Hypothetical protein	86	2.79	542	18.47

Plasmids	*4*				*1*
Name	pSR11	pSR56	pSR61	pSR84	pSR35
Length (Mb)	11.23	56.53	61.37	84.34	35.5
G+C content	63.29	60.03	59.58	63.19	59.70
ORF number	13	38	50	70	33
Total orthologous[a]	1	11	16	28	nd
COG categories[b]					
K-Transcription	0	1	0	2	0
L-DNA replication, recombination and repair	0	7	5	12	6
D-Cell division and chromosome partition	1	1	1	0	1
V-Defense mechanisms	0	4	0	0	1
M-Cell envelope biogenesis, outer membrane	0	0	1	12	0
N-Cell motility and secretion	0	0	0	2	0
U-Intracellular trafficking, secretion, and vesicular transport	0	0	1	0	0
T-Signal transduction mechanisms	0	0	0	7	0
H-Coenzyme metabolism	0	0	0	1	0
E-Amino acid transport and metabolism	0	1	0	0	0
R-General function prediction only	0	0	2	0	1
S-Function unknown	0	3	1	3	1
No COG categories					
Hypothetical proteins	9	14	30	25	23
Conserved hypothetical proteins	3	7	9	10	

Abbreviations: ANI, average nucleotide identity; COG, clusters of orthologous group; ORF, open reading frame; rRNA, ribosomal RNA; tRNAs, transfer RNAs.
[a]In all, 197 additional ORFs not previously annotated were likely orthologs of the corresponding M8 genes were detected in M31 genome.
[b]Many of the genes have assignment to more than one COG category.

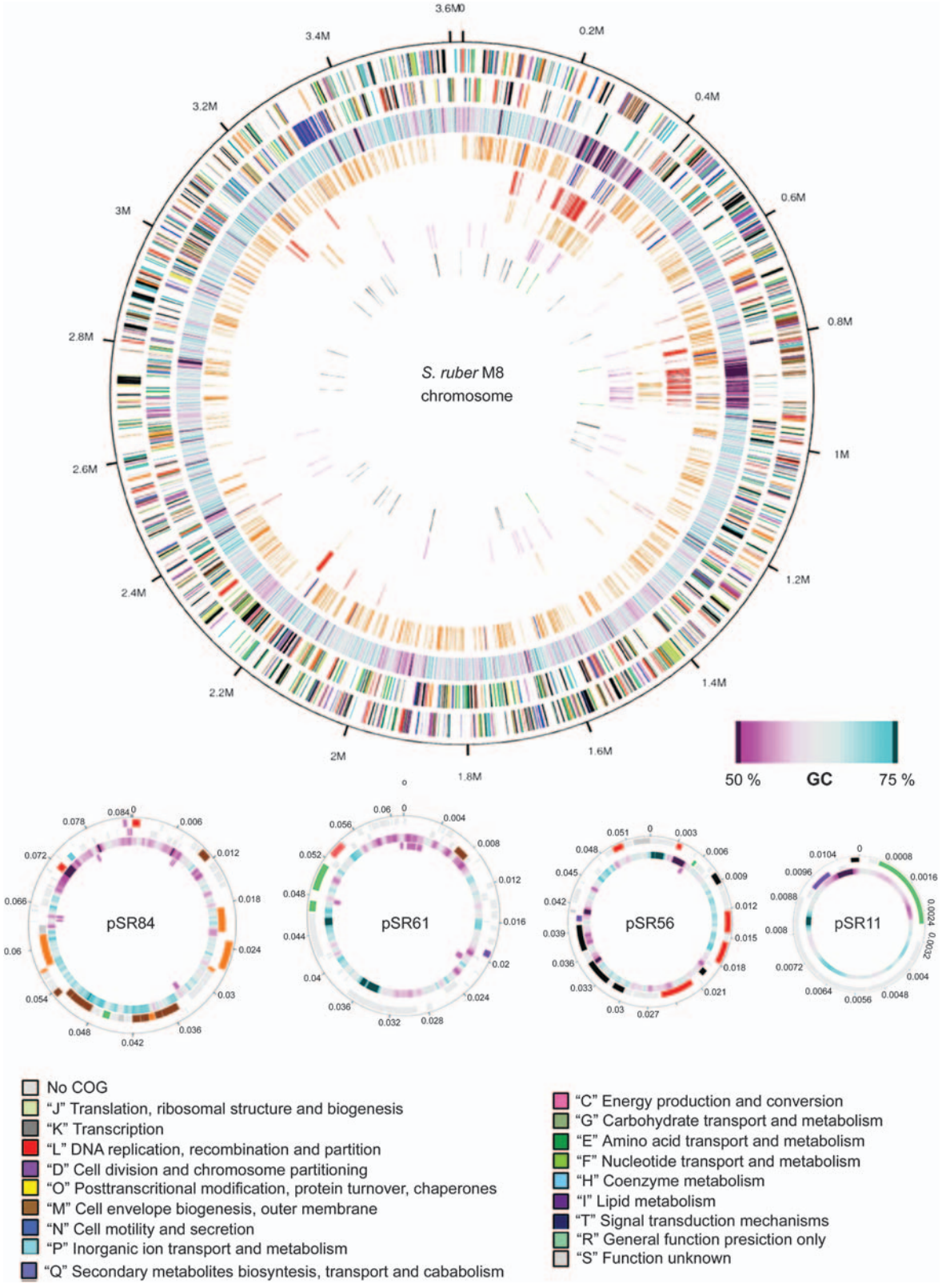

Figure 1 Circular representations of the M8 chromosome and its plasmids. For the chromosome, from the outside to inside: **Circles 1–2**: annotation by clusters of orthologous group (COG) functional categories (see insert). **Circle 3:** G + C content. **Circle 4**: divergent genes in *Salinibacter ruber* M8/M31 marked in blue and orange the genes below 50% or 90% identity, respectively. **Circle 5:** genes present in M8 and some other organisms in the database but absent in M31. **Circle 6:** strain-specific genes in *S. ruber* (genes absent from M31 and databases). **Circle 7:** transposases. **Circle 8**: putative LGT from *Archaea*. For the plasmids, the outer and second circles represented open reading frames (ORFs) on the plus and minus strands, colored by COG functional categories. The third circle represents the G + C content.

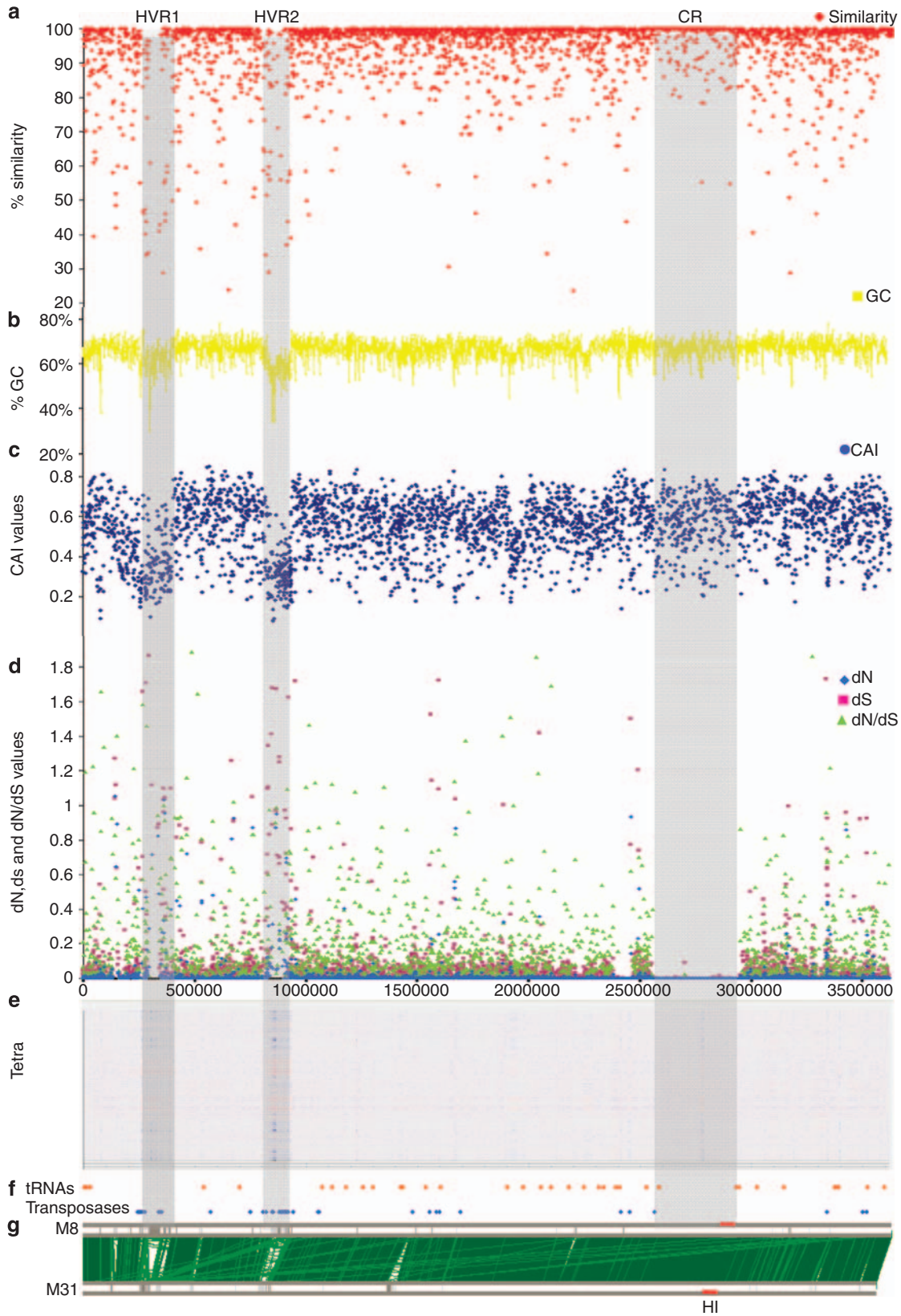

Figure 2 Summary and data comparison of *S. ruber* M8 chromosome. (**a**) Identities to M31 orthologous genes. (**b**) G+C content. (**c**) CAI indexes. (**d**) dN, dS and dN/dS values for homologous. (**e**) TETRA (tetranucleotide usage patterns in DNA sequences). (**f**) Transfer RNAs (tRNAs) and transposases along the chromosome. (**g**) Whole-genome alignment between *S. ruber* M8 and M31.The sequences have been aligned from the predicted replication origin. The green bars linking both genomes represent ortholog matches identified by FASTA (fast nucleotide comparison) analysis with a 100 nucleotide window.

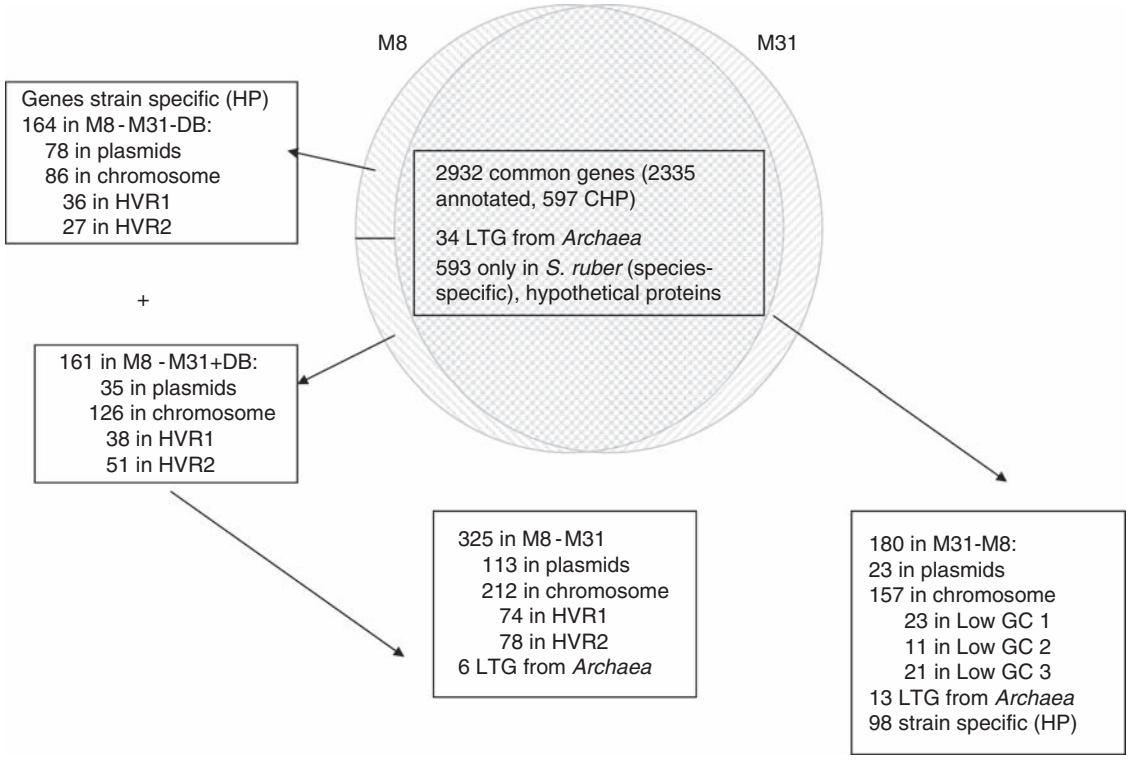

Figure 3 Overall comparison of M8 and M31.

than those based on the comparison of the full sequences (ANIb). The latter yields results that are more comparable to those obtained by DNA–DNA hybridization experiments. Low values of amino acid identities have also been reported for distinct ecotype representatives within other species such as *Prochlorococcus* spp. (Coleman *et al.*, 2006) and *A. macleodii* (Ivars-Martínez *et al.*, 2008).

Phylogenetic reconstructions revealed 20 gene duplications in the M8 genome that were absent from M31, and that likely occurred after the divergence of both strains. As many as 14 of these duplications corresponded to genes encoding putative transposases. Transposase duplication is sometimes associated with IS (insertion sequences) transposition events, although for most cases we did not detect other typical IS features such as terminal inverted repeats or flanking direct repeats.

The average non-synonymous (dN) to synonymous (dS) substitution rates for *S. ruber* genes was 0.125. This value lies between that of *Escherichia coli* (0.081) and that reported for pathogens inhabiting a narrow range of niches such as *Helicobacter pylori* (0.188) and *Neisseria meningitidis* (0.158) (Jordan *et al.*, 2002). The low value in *E. coli* is considered to be related to an enhanced purifying selection caused by its large population size (Jordan *et al.*, 2002). *S. ruber* would be an intermediate case, because its population sizes are remarkably high, but it does not seem to undergo dramatic niche changes. Forty-one genes had dN/dS values above one, which means that they are probably under positive selection (Jordan *et al.*, 2002), and 25 of these are species-specific hypothetical proteins. A possible explanation for these high ratios is that the genes were acquired recently and are still in the process of adaptation to their new context. Unusually high dN/dS ratios have also been observed in other closely related strains (Rocha *et al.*, 2006) as a result of paralog formation, that is, recent gene duplications in which one gene undergoes a process of neo- or sub-functionalization.

Remarkable is the presence of a region (2 583 792–2 960 363 bp, 'CR') with a high degree of sequence conservation (99.5%) in which there were no non-synonymous changes. In the cases in which the similarity among the predicted proteins was not 100%, the differences were always caused by insertion or deletions within the genes. This region was enriched in genes of the clusters of orthologous groups categories E (amino acid transport and metabolism) and P (inorganic ion transport and metabolism), with most genes belonging to the transport rather than metabolism part of these categories. More specifically, all the CR-encoded proteins in clusters of orthologous group class P were inorganic ion transporters (mostly for potassium, iron and phosphate). In addition, eight out of nine amino acid transporters that were annotated in the genome were located in the CR. In good agreement, rapid annotations using subsystems technology analyses (Aziz *et al.*, 2008) indicated that genes

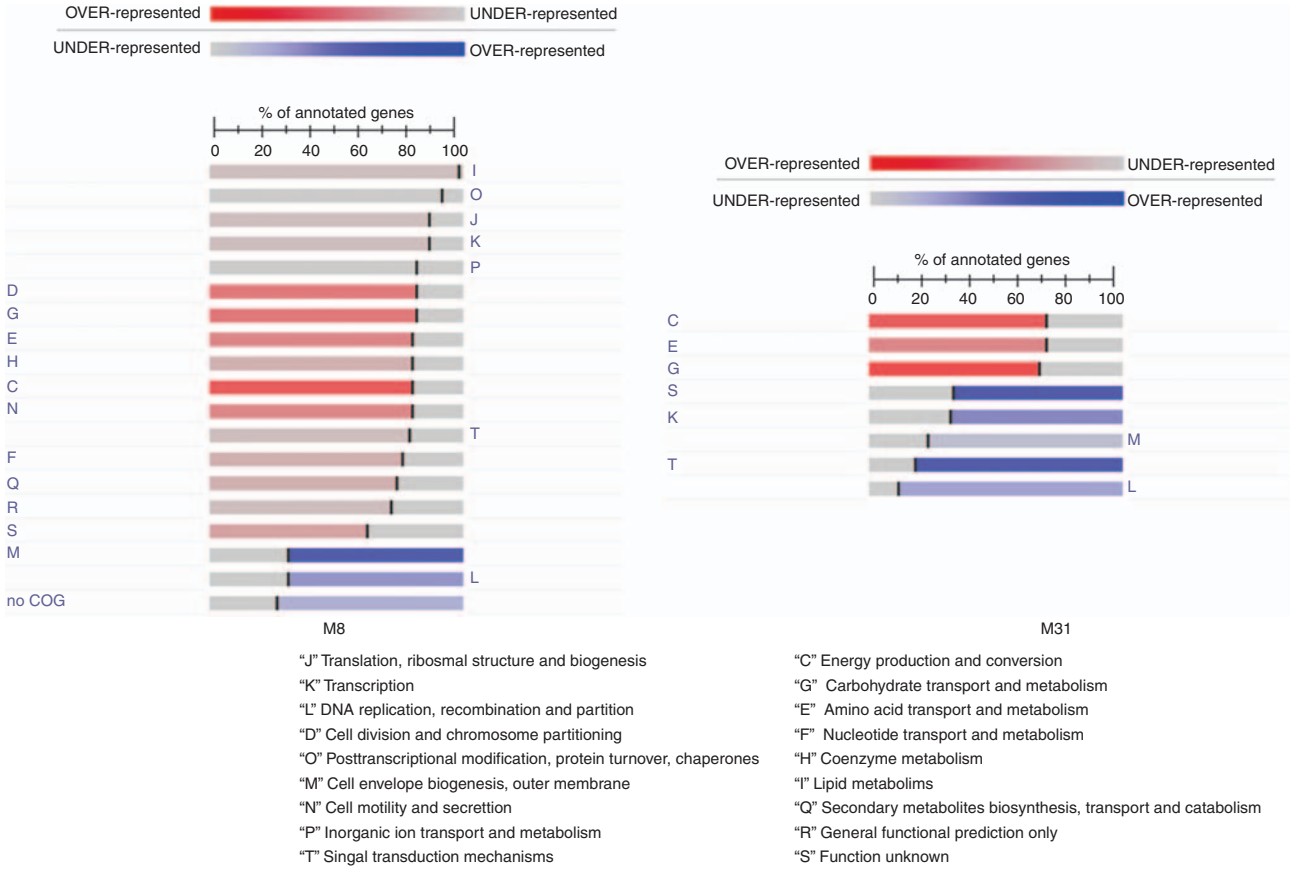

Figure 4 Over- and underrepresented clusters of orthologous group (COG) functional classes in M8 and M31 genes ranked according their codon usage. Classes with red bars are over-represented among the genes with lower CAI value, while classes with blue bars aver over-represented among genes with higher CAI values. Over- and underrepresentation was estimated by a segmentation test as implemented in the Fatiscan program.

of the subsystems 'potassium metabolism', 'membrane transport' and 'phosphorus metabolism' were highly enriched in the CR. These zone only contained genes present in both strains, including the transporters encoded in the 'halophilic island' that has previously been reported for *S. ruber* M31 (Mongodin *et al.*, 2005).

Codon usage indexes were rather constant along *S. ruber* genome, with the exception of HVRs in which genes with poorer indexes were concentrated. With only one exception, the genes with the 10% highest codon adaptation indexes (CAIs, a measure of codon usage) (above 0.716), were common to M8 and M31. However, M8 and M31 showed also small differences in codon usage (Figure 4). It is noteworthy that many genes involved in transport as well as in energy and nucleotide metabolism show the highest CAIs, while genes with low indices are mostly hypothetical proteins and transposases. A majority of the ribosomal proteins had an average CAI, which is typical for many slow-growing microorganisms (Carbone *et al.*, 2003) and thus applies to most environmental bacteria.

Hypervariable regions or genomic islands. Two large zones of anomalously low G+C content, poorer CAIs and deviant tetranucleotide signatures were identified in the chromosome of M8 at positions 0.25 and 0.82 Mb from the origin of replication. We will refer to these regions as the hypervariable regions HVR1 and HVR2 (as in Whilhelm *et al.* (2007)). HVRs corresponded (see Figure 1, bottom and Supplementary Figure 2) with two of the three genomic islands previously described in *S. ruber* M31 (Mongodin *et al.*, 2005) and could thus represent islands of high genomic variability in this species. A general terminology for hypervariable regions does not seem to exist, however, they are often referred to as 'islands' in analogy to pathogenicity islands, although genomic islands are usually considered to be mobile genetic elements (Dobrindt *et al.*, 2004). Genomic islands have been reported for many prokaryotes, ranging from human pathogens to free-living marine bacteria or extremely halophilic *Archaea*. However, in contrast to most of the regions described so far (Dobrindt *et al.*, 2004), HVRs in *S. ruber* M8 were clearly not associated with transfer RNAs (Figure 1). Such a lack of transfer RNA association has also been found for metagenomic islands detected in the environmental metagenome of the extremely halophilic

Table 2 General characteristics of the HVRs presents in *S. ruber* M8

	HVR1	*HVR2*
General characteristics		
Average GC%	60.6	58.2
Origin	256 919	830 124
End	396 085	956 929
Length (bp)/%	139 166/3.8	126 845/3.5
Coding density %	75.5	59.84
Genes	*Number/%*[a]	*Number/%*
M8–M31–DB[b]	36/41.86	27/31.40
M8–M31+DB[c]	38/30.16	51/40.48
M8+M31–DB[d]	15/2.56	14/2.56
M8+M31+DB[e]	42/1.83	59/2.58
Total:	131/4.25	151/4.89
Transposases	16/25.40	20/31.75
Glycosyltranferases	9/30	3/1
Sulfotransferases	3/37.5	2/25
COG M	13/8.18	22/13.84
Phage related	0/0	3/42.86
CHP	41/3.43	61/5.11
HP	36/41.86	27/31.40

Abbreviations: CHP, conserved hypothetical protein; COG, clusters of orthologous group; HP, hypothetical protein; HVRs, hypervariable regions.
[a]Percentage of the total genes in this category, in the chromosome.
[b]Genes M8 specific, absent from M31 and the rest of the database.
[c]Genes present in M8 and some other organisms in the database but absent from M31.
[d]Genes species specific, present only in M8 and M31.
[e]Genes present in M8, M31 and some other organism in the database.

archaeon *H. walsbyi* and was suggested as a characteristic of archaeal genomic islands (Cuadros-Orellana *et al.*, 2007). Annotation and functional classification revealed that 21% of the genes belonging to the clusters of orthologous group M (cell envelope biogenesis, outer membrane) are located in HVRs, which account only for 7.7% of the chromosome. This difference in gene content is particularly pronounced within HVR2, for which rapid annotations using subsystems technology analyses (Aziz *et al.*, 2008) only detected genes belonging to the subsystem 'cell wall and capsule'. Similarly, the presence of genes coding for transposases, glycosyltransferases, and sulfotransferases was found to be remarkably high in HVRs (Table 2).

When the M8 genome was used for contig recruitment of environmental clones from a solar saltern in San Diego, a lower coverage of HVR was found, indicating that these regions were underrepresented in the metagenome (Supplementary Figure 3). Comparisons of genomes from *Prochlorococcus* (Coleman *et al.*, 2006), *Haloquadratum* (Cuadros-Orellana *et al.*, 2007) and *Pelagibacter* (Whilhelm *et al.*, 2007) with corresponding metagenomes have provided similar results showing underrepresentation of HVRs in contig recruitment. Similarly, M8 strain-specific genes (Figure 2), extremely divergent orthologous genes (identities below 50%) (Figure 2, blue lines in the forth circle inward), and genes whit higher dNs values are overrepresented in these zones. Combined codonw and CAIJAVA results (Supplementray Table 1), that can provide valuable hints on the expression levels of genes because a high degree of codon adaptation results in higher translational efficiencies and thus higher expression levels, and their visualization (Figure 2) revealed that many genes within the HVRs have low CAI values, which could be an indication of recent acquisition.

*Involvement of interdomain (*Archaea-bacteria*) LGT in the shaping of* S. ruber *genome*

As *S. ruber* shares its habitat with extremely halophilic *Archaea*, in particular with *H. walsbyi*, it has been suggested as a good candidate for interdomain LGT events. This hypothesis was reinforced by the many phenotypic similarities found between *S. ruber* and many extremely halophilic *Archaea* (Antón *et al.*, 2008). The analysis of the *S. ruber* M31 genome suggested that this was indeed the case, although the number of genes likely involved in LGT was found to be lower than initially anticipated (Mongodin *et al.*, 2005). To increase the stringency of these analyses, we used a two-step approach to identify genes likely transferred to the M8 genome from *Archaea* (or the other way around). In the first step, we identified 40 candidate genes for interdomain LGT (Supplementary Table 2) by sequence comparison and phylogenetic analysis, which were then scrutinized by additional methods in the second step. A majority of these genes coded for membrane proteins, including retinal-binding proteins, adaptive-response sensory kinases and ion uptake proteins. Moreover, 6 of the 40 LGT candidates were strain-specific for *S. ruber* M8 and, conversely, 13 of the 47 genes that were reported earlier as inter-domain LGT candidates in *S. ruber* M31 (Mongodin *et al.*, 2005) lacked any homolog in *S. ruber* M8.

We subsequently analyzed LGT candidates using self-organizing maps and also compared the di, tri- and tetranucleotide frequencies found in the 40 LGT candidate genes to the corresponding frequencies within 572 prokaryote genomes and plasmids (Supplementary Table 2). Many of them exhibited weak similarities to halophilic *Archaea*, most pronouncedly reflected in their dinucleotide signature. However, for the majority of the genes the results lacked consistency over all oligonucleotides and therefore the derived classification was ambiguous. In some cases, *S. ruber* oligonucleotide frequencies also showed good correlations to those of *Actinobacteria*, most likely a result of their common high G+C contents. The results of the self-organizing map-classification are summarized in Supplementary Table 2. A majority of the LGT candidates were classified to *S. ruber* and only eight LGT candidates were found to have a clear archaeal signal according to this procedure. Three of these

genes (Supplementary Table 2) were present in *S. ruber* M8 but not in *S. ruber* M31.

As the presence of genes with an atypical codon composition is a strong indicator of LGT (Abby and Daubin, 2007), we performed a codon usage analysis of the LGT candidates. Codon usage analysis of the LTG candidates revealed average to good codon usage adaptation indices for most of the LGT candidates present in both strains. This might be an indication that in these cases, the LGT was not recent and the codon usage of these genes has been adapted to the *S. ruber* host genome. Consistently, LGT candidates present in M8 but not in M31 had rather poor codon usage indices, suggesting more recent LGT events, most likely after the separation of the M8 and M31 lineages.

Reverse transcriptase-PCR analyses indicated that 31 of the 34 inter-domain LGT candidates present in both strains were transcribed at some point of the growth curve, although the expression patterns of some genes showed slight differences between the two strains. The residual three genes were not expressed in any strain under any of the conditions analyzed. Under standard conditions, expression could only be detected for three out of the six *S. ruber* M8-specific inter-domain LGT candidates (Supplementary Table 2, blue color). Similarly, only one of the eight *S. ruber* M31-specific inter-domain LGT candidates analyzed was found to be expressed under standard growth conditions. Overall, these results indicate that most of the species-specific inter-domain LGT candidates represent functional genes and are expressed during normal growth under laboratory conditions, although for the strain-specific inter-domain LGT candidates this is only the case for a small fraction.

Thus, besides being constitutively transcribed, most LGT candidates are not clustered in HVRs, instead they constitute part of the (so far) core genome of the species, including the halophilism island present in CR. In a way, LGT from *Archaea*, although not very extensive, has had a role in the shaping of *S. ruber* species.

Metabolomic comparisons of M8 and M31

As one of the most apparent differences in HVRs is the abundance of genes coding for sulfotransferases and glycosyltransferases, the metabolomes of M8 and M31 were compared for the presence of sulfonated/S-containing and glycosylated metabolites.

FT-ICMS comparison of signal pairs at exact mass differences corresponding to sulfonation and glycosylation indicated a consistent increase in sulfonated and glycosylated metabolites in M8 compared with M31 and this increase was considerably higher in the extracellular fraction of the cultures (Table 3). In addition, the exact masses obtained with electrospray ionization (−)-ultrahigh resolution mass spectrometry were converted into more than 2000 elementary compositions (C, H, O, N and S).

Table 3 Proportion of sulfonated and glycoylated metabolites in M8 and M31

	Salinibacter ruber	
Metabolites	*M8*	*M31 DSM 13588*
All fractions[a]		
Sulfonated	1.7	1.5
Glycosylated	4.4	3.1
Extracellular fraction		
Sulfonated	2.7	1.8
Glycosylated	6.1	2.6

[a]Intracellular soluble, pellets and extracellular (see Materials and methods).

These analyses also supported a significant increase of sulfur containing metabolites mainly in the extracellular fraction of M8 versus M31 from 21.8% to 28.5% relative to all calculated CHO (Carbon, Hydrogen, Oxygen) type of molecules.

Finally, the exact mass lists were assigned to specific metabolites of *S. ruber* within the KEGG database using the MassTRIX annotation interface (Suhre and Schmitt-Kopplin, 2008, www.masstrix.org); with this approach only 15–25% of the masses are assigned to metabolites in possible pathways. Small differences were observed between M31 and M8 metabolomes. However, cytoplasmic and extracellular fractions of M8 contained more metabolites related to pathways involving amino acids and carbohydrates/fatty acids, than the equivalent fractions in M31. When the Japanese Metabolome database (www.metabolome.jp) was used for metabolite annotation, a significant increase to more than 26% was found with respect to the number of S-containing molecules in the extracellular M8 fraction, such as hexose-sulfates, gluthathiones (see Figure 5), sulfobenzoate or sulfobenzaldehyde. When comparing for glycosylated metabolites, a systematically higher number of hits was found also for M8 when compared with M31. Regardless of the database used for metabolite annotation, the proportion of hits within supernatant samples to glycosylated metabolites relative to all hits is significant in M8, (+10% to +60% from the Japanese database and from +10% to +180% from the MassTRIX approach for cytoplasmic and extracelular fractions respectively).

Thus, the main metabolomic differences detected by all approaches in M8 relative to M31 are related to molecules released to the medium or loosely attached to the cell surface that could have been released during the sample processing.

Phage susceptibility

Many of the M8 and M31 (Mongodin *et al*., 2005) genes in HVRs were related to cell-surface properties that could have a role in phage recognition and evasion. Metabolomic analyses also indicated

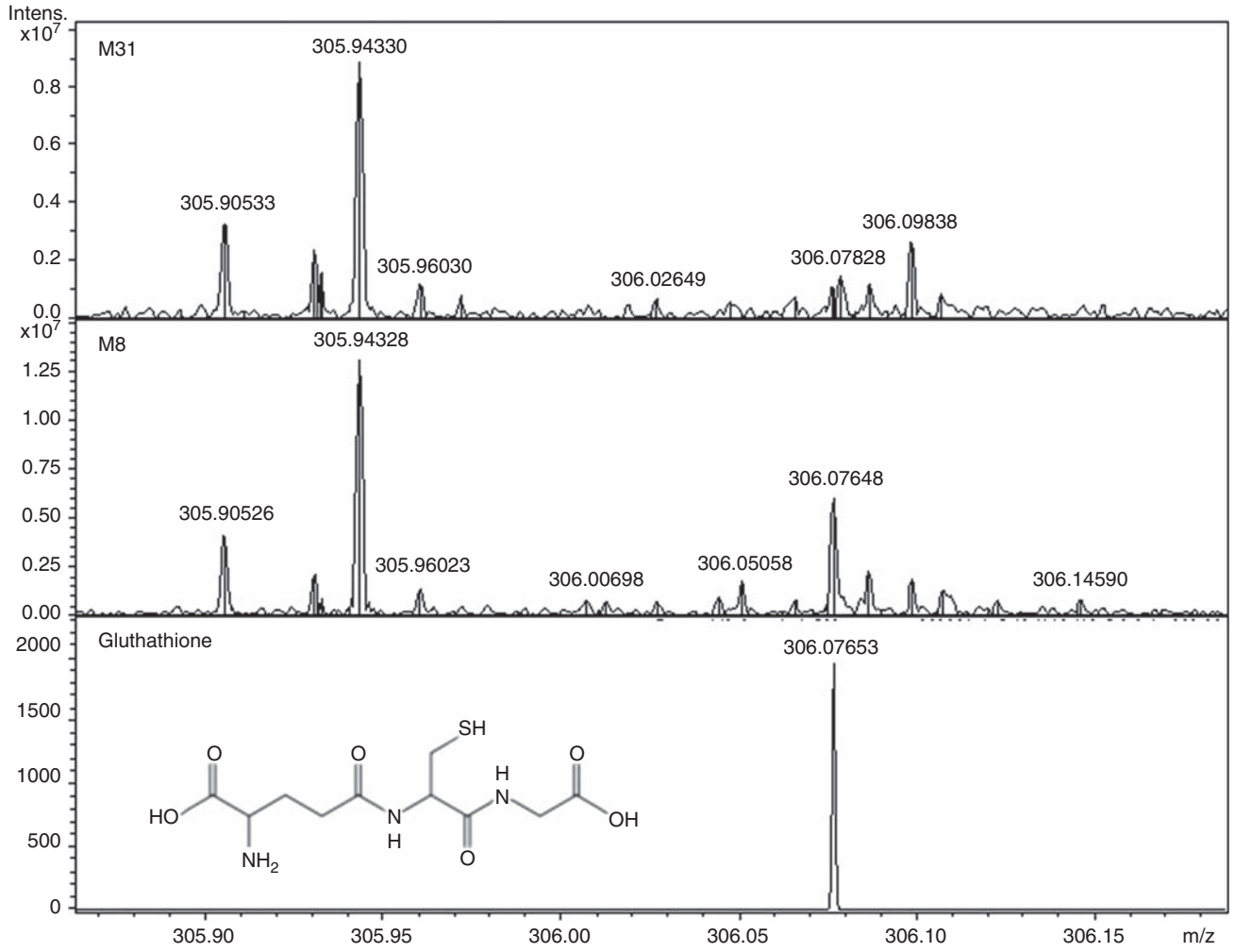

Figure 5 Example of relative increase in ion (C10H16O6N3S)—in supernatant of M8 relative to M31, corresponding to the elementary composition of gluthathione (simulated spectra).

differences in the exposed components of M8 and M31. Hypersaline environments show one of the highest numbers of viruses reported for planktonic systems with values of up to 2×10^9 virus-like particles per ml in crystallizers from which *Salinibacter* representatives have been isolated (Guixa-Boixareu *et al*., 1996). Halophages are thus considered to be important agents in mortality in hypersaline environments, given that no bacteriovory can be detected at salinities above 25% (Guixa-Boixareu *et al*., 1996). Therefore, cell-surface properties are likely under strong selection, as suggested for the extremely halophilic archaeon *H. walsbyi* (Cuadros-Orellana *et al*., 2007), and for marine bacteria that are exposed to high viral predation, for example, *Prochlorococcus* (Coleman *et al*., 2006) or *Pelagibacter* representatives (Whilhelm *et al*., 2007) for which surface-related proteins have also been observed in genomic island(s) or hypervariable regions within their genomes. Phages thus seem to have a high effect in driving micro-diversification within prokaryote species (Rodríguez-Valera *et al*., 2009), not only because of their direct selective pressure on exposed cellular components but also because of their role in LGT processes. Besides cell-surface manipulation and exopolysaccharide production, *Bacteria* and *Archaea* have other viral defense mechanisms such as the CRISPR (clustered regularly interspaced short palindromic repeats) system that has been proved to confer resistance to phage infection in *Streptococcus thermophilus* and, according to a wealth of metagenomic data, are considered to be a general mechanism of phage defense (Wilmes *et al*., 2008). However, *S. ruber* M8 and M31 lack this system.

According to the genomic and metabolomic data, it should be expected than strains M8 and M31 would show different phage susceptibility. Indeed, when we first tried to isolate *S. ruber* phages from Mallorca sample waters, we only succeeded in getting phages from M31, while not a single phage from M8 could be obtained. To quantify the different levels of phage susceptibility in both strains, we

infected cultures with virus assemblages from two different water samples (salt concentrations of 23.2% and 34.2%, see Materials and methods) obtained from Santa Pola salterns. For the lower salt water, M8 had plaque counts of $4.47 \times 10^2 \pm 61.1$ PFU ml^{-1} of natural sample and M31, $6.33 \times 10^2 \pm 65.1$ PFU ml^{-1}. However, for the high salt sample, M8 had a higher plaque count ($1.6 \times 10^4 \pm 3.61 \times 10^3$ PFU ml^{-1}) while no plaques were formed by M31 under the assayed conditions.

Competition between strains M8 and M31 under saturated salt conditions

The observation of metabolome differences between *S. ruber* strains M8 and M31 prompted us to analyze whether one of the two strains would have a growth advantage over the other under certain conditions. Thus, we set up competition experiments in which mixed cultures of *S. ruber* M8 and M31 were compared with pure cultures of each strain. Growth was monitored by OD (optical density)-reading and 4′,6-diamidino-2-phenylindole (DAPI) counting, and in mixed cultures cell numbers for M8 and M31 were determined also by quantitative PCR (see Supplementary Material). We tested for possible competition between the two strains under standard growth conditions and under salt-saturated conditions as they occur also in the natural habitat because crystallizer ponds are frequently saturated in NaCl. Under standard growth conditions, *S. ruber* M8 outcompeted *S. ruber* M31 (Figure 6a). However, in salt-saturated medium (Figure 6b), the density of *S. ruber* M31 in mixed cultures was roughly up to 30-fold higher than that of *S. ruber* M8, although in pure cultures the difference was only twofold. This indicates that under saturated salt conditions *S. ruber* M31 hinders growth of *S. ruber* M8, which in turn strongly suggests direct competition between the two closely related populations also *in situ*. Indeed, the salterns are periodically submitted to cycles of refilling and emptying, and thus *Salinibacter* spp. are exposed to fluctuating salt concentrations. This does not necessarily mean that these two strains are actually competing in nature, because dilution of every genotype could be too high to allow for the competition between them, such as reported for *V. splendidus* (Thomson *et al.*, 2005). As pointed out by Wilmes *et al.* (2008), 'the set of variable genes and genome rearrangements may be so large in some populations that no individuals have exactly the same genotype'. In the case of *S. ruber*, our previous studies indicate a high degree of genomic diversity between different isolates from the same environment (unpublished results) although, in any case, we show here that there is a real possibility of competition of extremely closely related strains. Then, given the appropriate selective force, one strain could displace the other and thus give raise to sympatric differentiation and proceed to broader genomic differences.

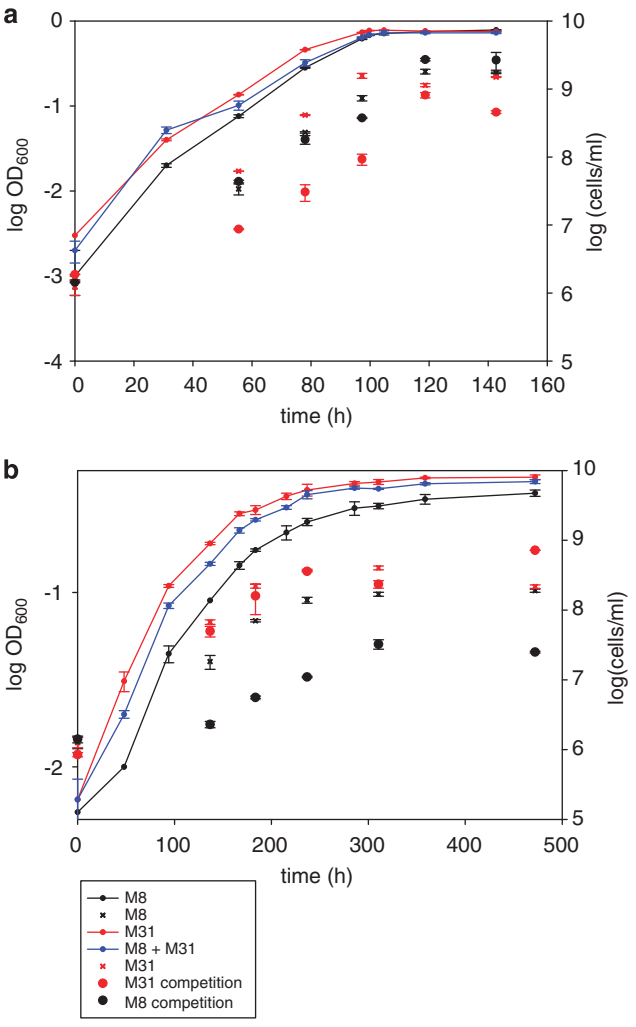

Figure 6 Competition experiments between strains M8 and M31 grown under standard conditions (**a**) and in NaCl-saturated medium (**b**). Lines represent OD reads for pure and mixed cultures, as indicated in the insert. For mixed culture, the values of M8 and M31 measured by quantitative PCR are represented by black and red dots, respectively. Crosses represent the values of M8 (black) and M31 (red) measured by fluorescence *in situ* hybridization (FISH) in pure cultures of every strain.

These changes in *S. ruber* strain abundances could never have been detected in nature by means of the commonly used rRNA-based standard tools of molecular ecology. Certainly, 16S rRNA genes are very conserved, and thus not useful to unveiling micro-diversity. However, ITS are far less conserved and frequently used for assessing coexisting populations of the same species *in situ* (for example, *Prochlorococcus*, (Coleman *et al.*, 2006)). In addition, many micro-diversity landscapes have been described in terms of rRNA diversity (Acinas *et al.*, 2004; Pedrós-Alió, 2006). Thus, the use of rRNA/ITS methods to detect population dynamics in natural samples provides only a partial picture because it can overlook ecological relevant changes in micro-diversity within a species.

Conclusions

Speciation is the generation of permanently distinct clusters of closely related bacteria (Fraser *et al.*, 2007). M8 and M31 are genetically distinct strains that, as shown by the competition and phage infection experiments, can respond differently to external conditions and could thus undergo sympatric differentiation in their environment. Whether differences between M8 and M31 are neutral or not depends more on the environment than on the genomes themselves, as reported for *Vibrio* and *Roseobacter* members (Polz *et al.*, 2006) that had differences adaptive under ecological circumstances and neutral under another. Although M31 could displace M8 under the appropriate circumstances, this may not be occurring in nature because competition can be part of a dynamic process, with a succession of different genotypes of the population that change their concentrations with time, originating a reservoir of genetic diversity inside a sequence space that would constitute the species *S. ruber*. We have captured the evolution of this species in the motion but the direction to which it will proceed cannot be predicted. To clarify this point, it would be necessary to study the changes within the M8/M31 genomes over time in their true ecological context, which includes not only habitat resources and physical conditions but also phages and predators as well as the substances secreted by other organisms (Cohan and Koeppel 2008).

Beyond the controversy of how many microbial species exist (Pedrós-Alió, 2006), or even if they exist at all (Doolittle and Zhaxybayeva, 2009), a new 'unknown' is rising regarding the functional diversity and environmental adaptation of bacteria. This study is showing the extent of genome, phenotype and niche differentiation encoded by rather small inter-strain differences, a phenomenon that most likely is very widespread in the environment.

Acknowledgements

We warmly thank Carlos Pedrós-Alió and Johaness Sikorski for the critical reading of a previous version of the paper and their many helpful suggestions and discussion. This work was funded by projects CGL2006-12714-CO2-01 and 02 from de Spanish Ministry of Science (to JA and RRM). J Dopazo thanks the National Institute of Bioinformatics (www.inab.org), that is a platform of Genoma España. This work was financially supported by the Max Plank Society within the projects Marine Genomics and EnviTools.

References

Abby S, Daubin V. (2007). Comparative genomics and the evolution of prokaryotes. *Trends Microbiol* **15**: 135–141.

Acinas SG, Klepac-Ceraj V, Hunt DE, Pharino C, Ceraj I, Distel DL *et al.* (2004). Fine-scale phylogenetic architecture of a complex bacterial community. *Nature* **430**: 551–554.

Antón J, Oren A, Benlloch S, Rodríguez-Valera F, Amann R, Rosselló-Mora R. (2002). *Salinibacter ruber* gen. nov., sp. nov., a novel, extremely halophilic member of the Bacteria from saltern crystallizer ponds. *Int J Syst Evol Microbiol* **52**: 485–491.

Antón J, Peña A, Santos F, Martínez-García M, Schmitt-Kopplin P, Rosselló-Mora R. (2008). Distribution, abundance and diversity of the extremely halophilic bacterium *Salinibacter ruber*. *Saline Syst* **4**: 1–10.

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA *et al.* (2008). The RAST server: rapid annotations using subsystems technology. *BMC Genomics* **9**: 75.

Carbone A, Zinovyev A, Kepes F. (2003). Codon adaptation index as a measure of dominating codon bias. *Bioinformatics* **19**: 2005–2015.

Cohan FM, Koeppel AF. (2008). The origins of ecological diversity in prokaryotes. *Curr Biol* **18**: R1024–R1034.

Coleman ML, Sullivan MB, Martiny AC, Steglich C, Barry K, DeLong EF *et al.* (2006). Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* **311**: 1768–1770.

Cuadros-Orellana S, Martín-Cuadrado AB, Legault B, D'Auria G, Zhaxybayeva O, Papke RT *et al.* (2007). Genomic plasticity in prokaryotes: the case of the square haloarchaeon. *ISME J* **1**: 235–245.

Dobrindt U, Hochhut B, Hentschel U, Hacker J. (2004). Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Micro* **2**: 414–424.

Doolittle WF, Zhaxybayeva O. (2009). On the origin of prokaryotic species. *Genome Res* **19**: 744–756.

Feil EJ. (2004). Small change: keeping pace with microevolution. *Nat Rev Microbiol* **2**: 483–495.

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness AR, Bult CJ *et al.* (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**: 496–512.

Fraser C, Hanage WP, Spratt BG. (2007). Recombination and the nature of bacterial speciation. *Science* **315**: 476–480.

Guixa-Boixareu N, Calderón-Paz JI, Heldal M, Bratbak G, Pedrós-Alió C. (1996). Viral lysis and bacterivory as prokaryotic loss factors along a salinity gradient. *Aquat Microb Ecol* **11**: 215–227.

Huerta-Cepas J, Bueno A, Dopazo J, Gabaldón T. (2008). PhylomeDB: a database for genome-wide collections of genes phylogenies. *Nucl Acids Res* **36**: D491–D496.

Hunt ED, David LA, Gevers D, Preheim SP, Alm EJ, Polz MF. (2008). Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science* **320**: 1081–1085.

Huson DH, Bryant D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**: 254–267.

Ivars-Martínez E, Martín-Cuadrado AB, D'Auria G, Mira A, Ferriera S, Johnson J *et al.* (2008). Comparative genomics of two ecotypes of the marine planktonic copiotroph *Alteromonas macleodii* suggests alternative lifestyles associated with different kinds of particulate organic matter. *ISME J* **2**: 1194–1212.

Jordan IK, Rogozin IB, Wolf YI, Koonin EV. (2002). Microevolutionary genomics of bacteria. *Theor Popul Biol* **61**: 435–447.

Meddini D, Donati C, Tettelin H, Masignani V, Rappuoli R. (2005). The microbial pan-genome. *Curr Opin Genet Dev* **15**: 589–594.

Mongodin EF, Nelson KE, Daugherty S, DeBoy RT, Wister J, Khouri H *et al.* (2005). The genome of *Salinibacter ruber*: convergence and gene exchange among hyperhalophilic bacteria and archaea. *Proc Natl Acad Sci USA* **102**: 18147–18152.

Papke RT, Koening JE, Rodríguez-Valera F, Doolittle WF. (2004). Frequent recombination in a saltern population of *Halorubrum*. *Science* **306**: 1928–1929.

Pedrós-Alió C. (2006). Marine microbial diversity: can it be determined? *Trends Microbiol* **14**: 257–263.

Peña A, Valens-Vadell M, Santos F, Buczolits S, Antón J, Kämpfer P *et al.* (2005). Intraspecific comparative analysis of the species *Salinibacter ruber*. *Extremophiles* **9**: 151–161.

Polz MF, Hunt DE, Preheim SP, Weinreich DM. (2006). Patterns and mechanisms of genetic and phenotypic differentiation in marine microbes. *Phil Trans R Soc B* **361**: 2009–2021.

Rocha EPC, Smith JM, Hurst LD, Holden MTG, Cooper JE, Smith NH *et al.* (2006). Comparisons of dN/dS are time dependent for closely related bacterial genomes. *J Theor Biol* **239**: 226–235.

Rodríguez-Valera F, Martín-Cuadrado AB, Rodríguez-Brito B, Pasić L, Thingstad TF, Rohwer F *et al.* (2009). Explaining microbial population genomics through phage predation. *Nat Rev Microbiol* **7**: 828–836.

Rosselló-Mora R, Lucio M, Peña A, Brito-Echeverría J, López-López A, Valens-Vadell M *et al.* (2008). Metabolic evidence for biogeographic isolation of the extremophilic bacterium *Salinibacter ruber*. *ISME J* **2**: 242–253.

Sikorski J. (2008). Populations under microevolutionary scrutiny: what will we gain? *Arch Microbiol* **189**: 1–15.

Suhre K, Schmitt-Kopplin P. (2008). MassTRIX: mass translator into pathways. *Nucl Acids Res* **36**: W481–W484.

Thomson JR, Pacocha S, Pharino C, Klepac-Ceraj V, Hunt DE, Benoit J *et al.* (2005). Genotypic diversity within a natural coastal bacterioplankton population. *Science* **307**: 1311–1313.

Whilhelm LJ, Tripp HJ, Givan SC, Smith DP, Giovannoni SJ. (2007). Natural variation in SAR11 marine bacterioplankton genomes inferred from metagenomic data. *Biol Direct* **2**: 19.

Wilmes P, Simmons SL, Denev VJ, Banfield JF. (2008). The dynamic genetic repertoire of microbial communities. *FEMS Microbiol Rev* **33**: 109–132.

Supplementary Information accompanies the paper on The ISME Journal website (http://www.nature.com/ismej)